Topic-Graph Based Recommendation on Social Tagging Systems: a study on ResearchGate

Yuyun Chen*, Hang Dong†‡, Wei Wang‡
*International Business School Suzhou, Xi'an Jiaotong-Liverpool University, Suzhou, China
†Department of Computer Science, University of Liverpool, Liverpool, United Kingdom
‡Department of Computer Science and Software Engineering, Xi'an Jiaotong-Liverpool University, Suzhou, China
Yuyun.Chen14@student.xjtlu.edu.cn, HangDong@liverpool.ac.uk, Wei.Wang03@xjtlu.edu.cn

ABSTRACT
Social Tagging Systems (STSs), allowing users to annotate online resources with freely chosen key words, are an essential type of application in Web 2.0. Recommendation in STSs can prevent information overload and support users to locate relevant items for interaction. This article applies a Topic-Graph Based Recommendation approach. First, we discover semantics behind tags through topic inferencing with Latent Dirichlet Allocation (LDA). Second, we conduct Graph-Based Recommendation for tags and users. The approach is applied on a real-word representative data sample collected from the Academic Social Networking Site ResearchGate. The widely used Co-occurrence Based Graph Recommendation is implemented as a baseline approach. Our preliminary human evaluation shows that the Topic-Graph Based Recommendation can complement to the Co-occurrence baseline to provide more reliable results. Future studies are provided on leveraging further features and information for recommendation from researcher-generated social media data on a large scale.

CCS Concepts
• Information systems→Social tagging • Information systems →Social recommendation

Keywords
Social Tagging Systems; Data mining; Graph-based recommendation; Probabilistic Topic Models; Academic Social Networking Sites

1. INTRODUCTION
Social Tagging Systems (STSs) are an essential type of application in Web 2.0. STSs allow users to upload, annotate, and share resources with other users, and most importantly, to stimulate user participation through annotation of resources with freely chosen keywords, known as tags [1-2]. This annotation activity results in a *folksonomy*, a social classification of online resources, consisting of inter-related users, tags and resources [1], [3]. Formally, the data structure of *folksonomy* is written as a tuple: $F = (U, T, R, Y)$ where U (users), T (tags), and R (resources) are finite sets, and Y is a ternary relation between them, i.e., $Y \subseteq U \times T \times R$, called tag assignments [3-4].

Many Social Networking Sites have an STS module which adopts user-generated tags as a backbone to support Information Retrieval and Question & Answering (Q&A). For example, in ResearchGate[1], users can annotate themselves as with "Topics" and "Skills and expertise" to epitomize and present their research areas representing their own professional identities. Other examples include "skills" in professional networking sevice Linkedin, "topics" or "tags" in Q&A systems Quora[2], Zhihu[3] and StackOverflow [4]. Although it is recognised that tags can potentially improve content organisation, with the vast amount of users, resources, and tags available online, the efficient filtering of content to prevent information overload remains a challenging task [2]. In this case, a line of study is recommendation in STSs which includes user recommendation, resource recommendation and tag recommendation [2].

This paper intends to mine the social tagging data from ResearchGate and provide graph-based recommendations of tags and users. More specific objectives are described as (i) tag recommendation, recommending tags to describe users' academic identity on ResearchGate and (ii) user recommendation, recommending users with common research interests. These recommendations of relevant items could help enrich user profiles and facilitate social interaction among users. In addition, mining tagging data in academic domains helps to find out connections among different research fields, which can be used to measure scholarly communication.

Co-occurrence is a widely used approach to explore relations in data and have been well adopted to recommendation in STSs [5]. Tag co-occurrence can be generally defined when two tags are used to annotate the same resource regardless of users, or when two tags are used by the same users regardless of resources [6]. However, using co-occurrence to find tag clusters ignores semantics, meanings, or denotation of tags. To mitigate this issue, Latent Dirichlet Allocation (LDA) is exploited to discover hidden semantic structures among words in large archives of documents [7-8]. LDA is a Probabilistic Topic Model, which states how documents are generated from different words according to latent variables in a probabilistic graphical model [7-8]. In this article, we apply a Topic-Graph Based Recommendation which leverages LDA to construct weighted graphs that link tags or users for recommendation. The approach generally contains two steps: (i) topic inferencing for item representation and (ii) graph construction for recommendation. Our preliminary human

[1] https://www.researchgate.net/

[2] https://www.quora.com/

[3] https://www.zhihu.com/topic

[4] https://stackoverflow.com/tags

evaluation on a representative sample of the real-word ResearchGate data shows complementary results between the Topic-Graph Based Recommendation and the Co-occurrence Graph Based Recommendation.

The remaining of the paper is structured into four sections. First, we introduce the related work about tag semantic discovery in section 2. Then the Topic-Graph Based Recommendation method is described in section 3. In the section 4, we introduce the data collection process from ResearchGate, our implementation of the Topic-Graph Based Recommendation and the Co-occurrence Graph Based Recommendation as a baseline, with graph visualisation and preliminary human evaluation. Conclusion and future studies are summarised in section 5.

2. RELATED WORK

Recommendation in STS is highly related to the discovery of semantics in tags. In recent years, numerous studies have been conducted in order to associate semantics to tags in folksonomies. Garcia-Silva *et al.* summarised a unified process that consists a set of common activities in most of the semantics association process [6]. The process is composed of four stages, *data selection and cleaning*, *context identification*, *disambiguation* and *semantic identification*. Many investigated works in the survey [6] followed this unified process and most adapted data co-occurrence as a heuristic to discover the tag semantics. Mika adopted a graph-based approach using the co-occurrence of tags on a Del.icio.us[5] dataset and generated two lightweight ontologies [9]. Hammasaki *et al.* extended Mika's work to take the co-occurrence of users in folksonomies into account and tested on data collected from a Social Networking System for an academic conference [10]. Ginnakidou *et al.* leveraged both the co-occurrence and an external knowledge base to measure the similarity of tags on Flickr[6] for clustering [11]. Compared with general STSs such as de.icio.us and Flickr covering tag vocabularies in wide domains, academic STSs, such as Bibsonomy[7] and CiteULike[8] are more related to scholarly communication, where tag vocabularies are slower to be accumulated and thus much sparser and challenging to process [12-13]. Jäschke *et al.* designed an algorithm to mine the association of triples in data from Delicious as well as BibSonomy [14]. Bastian *et al.* also constructed a folksonomy of "skill and expertise" in LinkedIn and implemented a skill tag recommendation module [15]. Similar to LinkedIn as a professional identity management website, ResearchGate has recently received tremendous success in researchers' communities [16-17], but little attention has been paid on mining the tags on ResearchGate.

Different from the studies above, our research utilises the Probabilistic Topic Model, LDA, to discover semantics from the researcher-generated social tagging data in ResearchGate and generate graph-based recommendations of tags and users. We refer to the widely used co-occurrence method as a baseline and found complementarity between the proposed approach and the co-occurrence method.

[5] https://del.icio.us/

[6] https://www.flickr.com/

[7] http://www.bibsonomy.org/

[8] http://www.citeulike.org/

3. METHODOLOGY

In this section, we demonstrate the Topic-Graph based recommendation method in detail. The approach constitutes of two steps, (i) topic inferencing for tag representation and (ii) graph-based tag and user recommendation.

3.1 Topic Inferencing for Data Representation

The topic inferencing step is performed through applying the Latent Dirichlet Allocation (LDA). LDA constructs a probabilistic graphical model to globally simulate the generation process of documents, and therefore to infer the latent semantic structure of the documents, represented by a document-topic distribution θ and a topic-word distribution β [7-8]. The two key dependencies in the document generation process can be described as conditional probabilities below:

$p(Z_{d,n}|\theta_d)$, generating a topic index for each word in the document based on the document-topic distribution, and then

$p(W_{d,n}|\beta_k, Z_{d,n})$, generating a word based on its topic index and the topic-word distribution,

where W_d represent the observed words in document d, $W_{d,n}$ is the word n in document d, θ_d is the document-topic distribution for document d, β_k is the topic-word distribution for topic k, and $Z_{d,n}$ indicates the topic assignment for the word n in document d.

Additionally, LDA assumes that the prior distribution of the document-topic distribution is drawn from a symmetric Dirichlet distribution with the concentration parameter α and the prior distribution of a topic-word distribution satisfies a symmetric Dirichlet distribution with the concentration parameter η, as shown below.

$$\theta_d \sim \text{Dirichlet}(\alpha) \quad (1)$$

$$\beta_k \sim \text{Dirichlet}(\eta) \quad (2)$$

Combining the information above, the joint distribution for the observed documents could be written as [8]:

$$p(\beta_{1:K}, \theta_{1:D}, Z_{1:D}, W_{1:D}) = \prod_{i=1}^{K} p(\beta_i) \prod_{d=1}^{D} p(\theta_d) \left(\prod_{n=1}^{N} p(Z_{d,n}|\theta_d)\, p(W_{d,n}|\beta_{1:K}, Z_{d,n}) \right) \quad (3)$$

Gibbs Sampling is an effective method to infer LDA latent variables, the document-topic distribution θ_d and topic-word distribution $\beta_{1:K}$ [18-19]. It is a special case of Markov-Chain Monte Carlo (MCMC) that involves a suit of approximate iterative techniques to sample values from complex models.

A prerequisite of LDA is to specify the total number of topics K that the whole input document set covers. To determine the value of K, we use perplexity to measure how well a probability model predicts. A lower perplexity indicates better performance of the model. Perplexity here is defined as:

$$\text{perplexity}\ (S) = exp\left\{-\frac{\sum_{d \in S} \log P(W_d)}{\sum_{d \in S} N_d}\right\} \quad (4)$$

Where S is the document set in the testing data as tag lists; for each tag list d belong to the testing data S, W_d is the sequence of

tags in the tag list, and N_d is the number of tags in the tag list. As perplexity will change with different number of topics K, we are able to determine the optimal number of topics based on the perplexity value.

Regarding the Dirichlet concentration parameters α and η, Steyvers and Griffiths [20] suggest that the two hyperparameters should depend on the number of topics and the vocabulary size. It is found $\alpha = 50/K$ (K is the number of topics) and $\eta = 0.01$ works well with many different text collections. We follow this setting in the experiment.

Thus, we represent tags as a probability distribution of the hidden topics after the inferencing process. With the Bayes' theorem, we can transform the $p(\mathrm{w}|z)$, corresponding to β obtained from the Gibbs Sampling, into the tag-topic distribution $p(\mathrm{z}|w)$, as shown in the Equation (5) below according to [21]. The $p(\mathrm{z}|w)$ is regarded as a tag vector representation and is used for generating a social graph of elements (tags, resources) in the tagging data for recommendation.

$$p(\mathrm{z}|w) \propto p(\mathrm{w}|z)p(z) \quad (5)$$

More directly, resources can be represented as the $p(\mathrm{z}|d)$, corresponding to the document-topic distribution θ obtained from the topic inferencing process.

The following part in this section introduces the graph-based tag and user recommendation process, which includes generating a similarity matrix from the topic-based representation and constructing a social graph of tags and resources for recommendation.

3.2 Graph-based Tag and User Recommendation

3.2.1 Similarity Matrix Generation

The *tag similarity matrix* $M_{tag} \in R^{|T|*|T|}$ is necessary for the generation of a social graph, where the dimension $|T|$ is the number of distinct tags in the data. Each element in the matrix M_{tag} is the similarity value between two tags. Cosine similarity is used as the similarity measure between vectors representations of two tags for its popularity in Information Retrieval [22], where vector representations of tags are obtained using the Equation (5).

The generation process of the *user similarity matrix* $M_{user} \in R^{|U|*|U|}$ is similar to the *tag similarity matrix* M_{tag} above.

3.2.2 Social Graph Generation

For graph-based recommendation and visualization, the *tag similarity matrix* and the *user similarity matrix* are converted to undirected weighted graphs, called *tag social graph* and *user social graph* respectively. Each node in the *tag social graph* is a tag in the distinct tag set T. Each edge corresponds to a similarity relation between two tags. Weights of edges are the corresponding similarity score in the *tag similarity matrix*. Analogously for the *user social graph*, the set of nodes belongs to the user set U and two nodes are connected by an edge whose weight is from the *user similarity matrix*.

A similarity threshold *TH* is set to filter out the tag-tag or user-user relation which has low similarity strength. We empirically set *TH* as 0.6 for tags and *TH* as 0.75 for users to retain a considerable number of nodes in the two social graphs.

The recommended item (tag or user) of an item I can be retrieved by selecting the neighbours of corresponding node of I in the social graph. The selected neighbours are then ranked by their edge weights to the node.

4. EXPERIMENTS: RECOMMENDATION ON RESEARCHGATE

This section describes the experimental setting of the recommendation methods using ResearchGate as a case study. First, we describe the data collection process. Then, we discuss implementation of the baseline approach, Co-occurrence Graph Based Recommendation, and the Topic-Graph Based approach for tag and user recommendation. Our comparison of the two approaches through a preliminary human evaluation suggests their complementarity in recommending tags and users.

4.1 Data Collection

ResearchGate has caught much attention in research communities as an Academic Social Networking Site [16-17], [23]. In this study we focus on the STS module in ResearchGate. To note that in ResearchGate, users annotate the "Skills and expertise" and "Topics" for themselves. In this scenario, the Folksonomies $F = (U, T, R, Y)$ is simplified to $F = (U, T, Y)$ as the user set U is same as the resource set R. The tag (or user) recommendation problem is to suggest new tags (or users) in the tag set T (or user set U) according to a known tag (or user).

We selected 6583 users from 9 disciplines/departments of different natures in the top-5 US universities based on the Total RG score in ResearchGate by the March of 2016, inspired by the data collection process in the study [24]. The Scrapy package[9] in Python is used to crawl users' identity information anonymously along with their tags, which include both "Skills and expertise" and "Topics" in users' profiles on ResearchGate. The data collection process is anonymous: each user name is replaced to an ID number. The input file combines each user's ID and his/her tags in a single line. After deleting the users who entered less than 3 tags, there remain 4794 users who collectively contributed 53737 tag annotations with 7259 distinct tags. The average number of tag annotations per user is about 11.2(≈53737/4794).

4.2 Co-occurrence Graph Based Recommendation

The Co-occurrence Graph Based Recommendation method is proposed as the baseline. The Tag (or User) Co-occurrence Graph is constructed as an undirected weighted graph, where each node is a tag (or user) and each edge indicates the normalised co-occurrence score between two nodes. The co-occurrence score $c_{a,b}$ for tag t_a and t_b is calculated as the number of users who annotated both tags. For graph generation, we normalised all the co-occurrence scores globally and set them as edge weights, according to the equation $h_{a,b} = c_{a,b}/\max(c)$, where $h_{a,b}$ indicates the weight of the edge linking tag t_a and t_b, $\max(c)$ returns the global maximum co-occurrence score. For comparison, the approach also generates recommendations as neighbours of an item (tag or user) in the graph. The Figure 1 below is an excerpt of the tag co-occurrence graph containing the user generated tag "computer engineering" on ResearchGate. Tags with high co-occurrence score such as "electrical engineering", "software engineering", "electronic engineering" and "c" are recommended.

[9] https://scrapy.org/

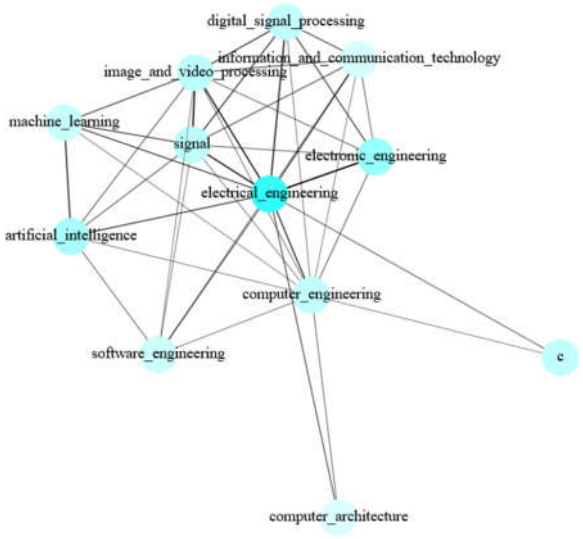

Figure 1. An excerpt of tag co-occurrence graph centring the user-generated tag "computer engineering" on ResearchGate, the recommended tags are neighbours of the node "computer_engineering" in the graph.

4.3 Topic-Graph Based Recommendation

We conducted the Topic-Graph Based Recommendation proposed in section 3. The topic inferencing with LDA is implemented using the MALLET Library[10]. We experimented the number of topics K from 10 to 150 with a step as 5 for increment. The perplexity result calculated from the Equation (4) is presented in the Figure 2 below.

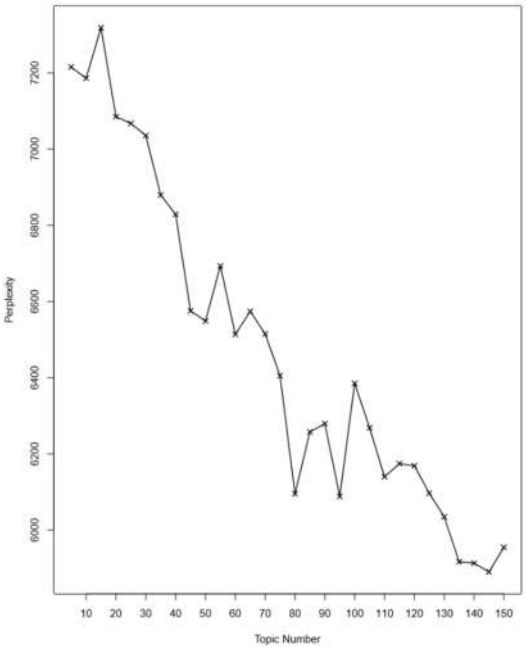

Figure 2. Perplexity change with respect to the number of topics K: The perplexity value (Y-axis) dropped in general when the number of topics K (X-axis) increased from 10 to 150.

[10] http://mallet.cs.umass.edu/topics.php

It seems overall the perplexity decreases when the number of topics K increases, which may suggest that a larger K is better. However, overfitting can occur if K is too large and we can see that perplexity still increase when K changes from 80 to 110. We therefore selected K as 80. The values of α and η are set as $0.625(= 50/K)$ and 0.01 respectively.

We generated the *tag social graph* and *user social graph* after the topic inferencing step. For tag recommendation, a proportion of the whole *tag social graph* is presented in the Figure 3, which focuses on the tag "computer engineering". Edges with similarity above 0.6 are selected. Recommended tags, including "asic", "vlsi", "software engineering", "image and video processing", are the linked neighbours of the provided tag "computer engineering" in the graph. Comparing the results from the Topic-Graph in Figure 3 with Co-occurrence graph Figure 1, we may find the Topic-Graph can complement to the Co-occurrence Graph for recommendation.

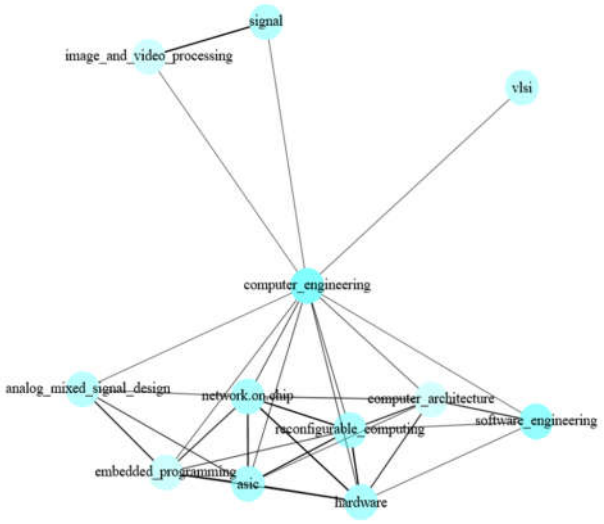

Figure 3. An excerpt of *tag social graph* centring the user-generated tag "computer engineering" on ResearchGate, the recommended tags are neighbours of the node "computer_engineering" in the graph. Compared to Figure 1, it seems the recommended tags are complimentary between the Topic-Graph and Co-occurrence Graph approaches.

For user recommendation, an example is given by focusing on the user "User 138" in the *user social graph*. Edges with the similarity above 0.75 are plotted in the Figure 4 below, where four users "User 381", "User 2003", "User 42" and "User 2053" are recommended given "User 138".

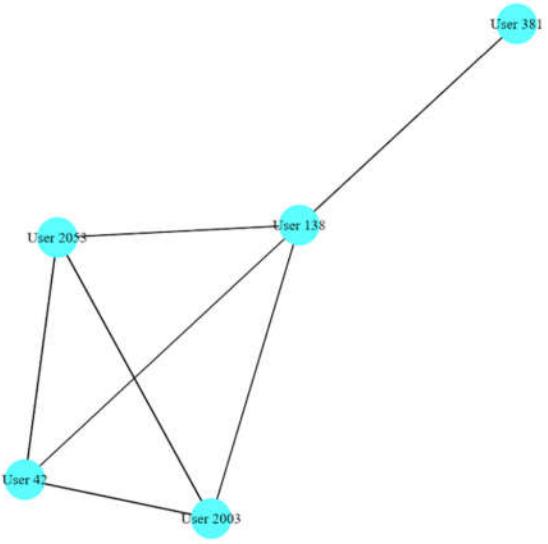

Figure 4. An excerpt of the *user social graph* centring the user "User 138" on ResearchGate, the recommended users are the linked neighbours of the node "User 138" in the graph.

To justify the user recommendation, we list the tags ("Skills and expertise" and "Topics") of these four recommended users as well as the input user in Table 1 below. It can be observed that the recommended users share semantically similar tags with the input user.

Table 1. Recommended Users to "User 138" and Their Tags

User	Tag
User 138 (input user)	Artificial Intelligence, Computer Science, Software Engineering, Cognitive Psychology, Data Mining, Programming Languages, Social Psychology, Human-Computer Interaction, Social Neuroscience, Social Psychology, Cognitive Science, Cognitive Psychology, Cognitive Neuroscience, Social Cognitive Neuroscience, Object-Oriented Programming, Artificial Intelligence, EEG/ERP, Mobile Application Development, Cross Cultural Psychology, Machine Learning, Operating Systems, Statistical Learning, Web Applications
User 381	Software Engineering, Programming Languages, Operating Systems, Computing, ASIC, Storage
User 2003	Cognitive Psychology, Human-Computer Interaction, Cognitive Science
User 42	Artificial Intelligence, Statistics, Cognitive Psychology, Human-Computer Interaction, Neuropsychology, Cognitive Science, Cognitive Neuropsychology, Memory, Working Memory, Cognitive Development, Cognitive Neuroscience
User 2053	Artificial Intelligence, Neuroscience, Cognitive Science

4.4 Preliminary Human Evaluation

Due to the relatively small amount of data collected for this study, automated evaluation based on holding out a part of data for testing does not seem reliable. For this preliminary evaluation, we adopt human evaluation which directly reflects the real-word scenario of the recommendation tasks. It is found that both the co-occurrence method and the LDA method provide reasonable results. There is also overlapping of tags recommended from the two approaches. At this stage, it is still not evident to decide which method is better than the other in terms of accuracy. We suggest that the tag recommendations from LDA could be used as a complement to the co-occurrence based method. This is probably because the topic inferencing in LDA can capture further semantics beyond the co-occurrence relation of tags.

To conclude from the preliminary evaluation based on human judgement, the Co-occurrence Graph Based Recommendation and the Topic-Graph Based Recommendation generated complementary results. We encourage a future study on the automated evaluation of the two approaches based on a large scale of user-generated data in Academic Social Networking Sites.

5. CONCLUSION AND FUTURE WORK

In this paper, we introduced Topic-Graph Based Recommendation approach and applied it on the STS module in an Academic Social Networking Site, ResearchGate. The tag recommendation and the user recommendation are conducted through the inference of LDA latent variables and the construction of social graphs. Based on a preliminary human judgement, it is found results of the two approaches are complementary in recommendation in our experiment.

One limitation of this study is that we only used a representative sample of data in ResearchGate in our experiment. It is worth to collect larger amount of data for further analysis. Also, the recommendation methods in this study could be expanded with richer features, including a thorough semantic identification to capture the synonym between tags, more advanced measure of similarity based on word embedding and other similarity metrics, and involving other information into the recommendation, such as users' online impact measures (RG scores) [17] and self-archiving behavioural data [24].

6. ACKNOWLEDGMENTS

This research is funded by the Research Development Fund at Xi'an Jiaotong-Liverpool University, contract number RDF-10-2015. Also, H. D. would like to thank J. Lee for his suggestions on preparing a representative sample dataset from ResearchGate when they conducted the research [24] together.

7. REFERENCES

[1] Vander Wal, T. 2007. *Folksonomy*. Retrieved from http://vanderwal.net/folksonomy.html

[2] Marinho, L. B., Hotho, A., Jäschke, R., Nanopoulos, A., Rendle, S., Schmidt-Thieme, L., Stumme, G., and Symeonidis, P. 2012. *Recommender Systems for Social Tagging Systems*. Springer New York, New York, NY.

[3] Singer, P., Niebler, T., Hotho, A., and Strohmaier, M. 2014. Folksonomies. In *Encyclopedia of Social Network Analysis and Mining*, R. Alhajj and J. Rokne, Ed. Springer New York, New York, NY, 542-547.

[4] Hotho, A., Jäschke, R., Schmitz, C., and Stumme, G. 2006. Information Retrieval in Folksonomies: Search and Ranking. In *The Semantic Web: Research and Applications: 3rd European Semantic Web Conference, ESWC 2006* Budva, Montenegro, June 11-14, 2006 Proceedings, Y. Sure and J. Domingue, Ed. Springer Berlin Heidelberg, Berlin, Heidelberg, 411-426.

[5] Belém, F. M., Almeida, J. M., Gonçalves, M. A. 2017. A survey on tag recommendation methods. *J. Assoc. Inf. Sci. Technol.* 68, 4, 830-844.

[6] García-Silva, A., Corcho, O., Alani, H., and Gómez-Pérez, A. 2012. Review of the state of the art: Discovering and associating semantics to tags in folksonomies. *Knowledge Engineering Review*, 27, 1, 57–85.

[7] Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. Latent dirichlet allocation. *the Journal of Machine Learning Research*. 3, 993-1022.

[8] Blei, D. M. 2012. Probabilistic topic models. *Commun. ACM*. 55, 4, 77-84.

[9] Mika, P. 2007. Ontologies are us: A unified model of social networks and semantics. *Web Semant. Sci. Serv. Agents World Wide Web*. 5, 5–15.

[10] Hamasaki, M., Matsuo, Y., Nisimura, T., and Takeda, H. 2007. Ontology extraction using social network. In *International workshop on semantic web for collaborative knowledge acquisition*.

[11] Giannakidou, E., Koutsonikola, V., Vakali, A., and Kompatsiaris, I. 2008. Co-clustering tags and social data sources. In *The 9th International Conference on Web-Age Information Management, WAIM 2008*, 317–324.

[12] Du, H., Chu, S. K. W., and Lam, F. T. Y. 2009. Social bookmarking and tagging behavior: an empirical analysis on delicious and connotea. In *Proceedings of the 2009 International Conference on Knowledge Management*.

[13] Dong, H., Wang, W., Coenen, F. 2017. Deriving dynamic knowledge from academic social tagging data: A novel research direction. In *iConference 2017 Proceedings*, 661–666.

[14] Jäschke, R., Hotho, A., Schmitz, C., Ganter, B., and Stumme, G. 2008. Discovering shared conceptualizations in folksonomies. *Web Semant. Sci. Serv. Agents World Wide Web*. 6, 1, 38–53.

[15] Bastian, M., Hayes, M., Vaughan, W., Shah, S., Skomoroch, P., Kim, H., Uryasev, S., and Lloyd, C. 2014. LinkedIn skills: large-scale topic extraction and inference. In *Proceedings of the 8th ACM Conference on Recommender systems*, Foster City, Silicon Valley, California, USA, 1-8.

[16] Van Noorden, R. 2014. Online collaboration: Scientists and the social network. *Nature*. 512, 7513, 126-129.

[17] Thelwall, M., Kousha, K. 2015. ResearchGate: Disseminating, communicating, and measuring Scholarship? *Journal of the Association for Information Science and Technology*. 66, 5, 876-889.

[18] Griffiths, T. L. and Steyvers, M. 2004. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101, suppl 1, 5228-5235.

[19] Heinrich, G. 2005. Parameter estimation for text analysis. Technical report.

[20] Steyvers, M. and Griffiths T. 2007. Probabilistic topic models. In *Handbook of latent semantic analysis*. 427, 7, 424–440.

[21] Griffiths T., and Steyvers M. 2003. Prediction and semantic association. In *Advances in neural information processing systems*, 11-18.

[22] Manning, C. D., Raghavan, P., and Schütze, H. 2009. *Introduction to information retrieval*. Cambridge University Press, New York, 121. Retrieved from https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf

[23] Williams, A. E. and Woodacre, M. A. 2016. The possibilities and perils of academic social networking sites. *Online Information Review*. 40, 2, 282-294.

[24] Lee, J., Oh, S., Dong, H., Wang, F., and Burnett, G. 2017. A framework for studying motivations for self-archiving on academic social network sites. In *iConference 2017 Proceedings*, 684-687.